

ESTUDO COMPARATIVO DE BUSCA E RECUPERAÇÃO DE PATENTES DA UFRGS INDEXADAS NA DII, ORBIT, INPI E GOOGLE PATENTS

Ana Paula Medeiros Magnus

Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil
magnus.ana@gmail.com

Fernanda Bochi

Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil
nandabochi@gmail.com

Leticia Angheben El Ammmar Consoni

Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil
le.angheben@gmail.com

Rene Faustino Gabriel Junior

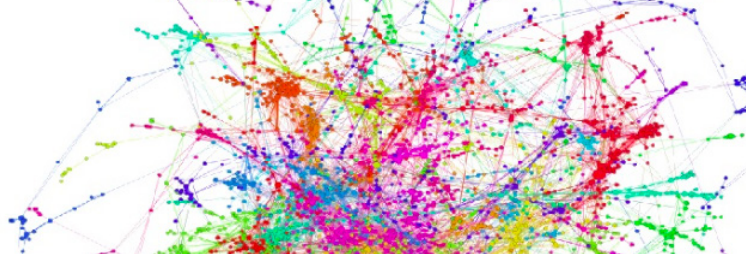
Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil
renefgj@gmail.com

Ana Maria Mielniczuk de Moura

Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil
ana.mmoura@uol.com.br

1 INTRODUÇÃO

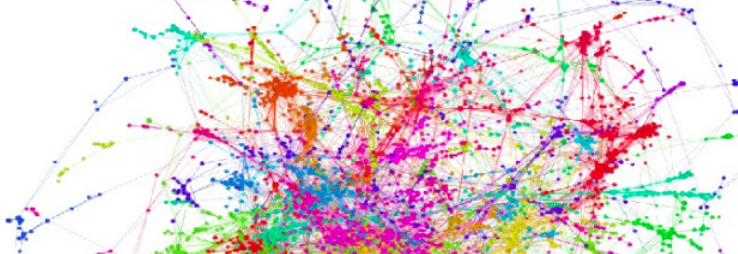
Uma das principais características da sociedade atual é a capacidade de produzir informações numa velocidade imensurável. Os avanços científicos e tecnológicos estão sendo desenvolvidos mais rápido que a capacidade das pessoas de processarem tais informações. De acordo com Caldas e Silva (2016), por um tempo a produção de dados científicos nada mais era do que dados com medidas entre *bit* e *byte*. Hoje se fala em *exabytes* de dados, que por sua vez são gerados diariamente, chegando a acumular no período de um ano quase dois *zettabytes* de dados. Números que crescem a cada instante com o avanço tecnológico, provocando com isso, uma necessidade de compreensão dos instrumentos que pro-



movem a busca, a coleta e a análise dos dados pesquisados. Com base nesse panorama, e sabendo que como fonte de pesquisas de patentes existe várias bases de dados disponíveis, ainda se desconhece suas coberturas, formas de recuperação, indexação e padronização dos dados para estudos patentométricos. Sendo assim, essa pesquisa tem como problematização saber: quais as características inerentes ao processo de coleta e recuperação de patentes depositadas nas bases *Derwent Innovations Index* (DII), *Orbit Intelligence* (Orbit), Instituto Nacional de Propriedade Industrial (INPI) e *Google Patents*. Tem como objetivo realizar um estudo comparativo nas bases já mencionadas, procurando identificar as características dos recursos e dos mecanismos de busca, aferir a cobertura, indexação, recuperação e padronização nas bases estudadas e comparar os resultados obtidos nelas.

2 METODOLOGIA

A presente pesquisa trata-se de um estudo descritivo, visando comparar as características de cobertura, indexação, recuperação e padronização das bases de dados DII, Orbit, INPI e *Google Patents*. Para efetivar a coleta de dados, realizada durante o mês de janeiro de 2018, e possibilitar as comparações utilizou-se a Universidade Federal do Rio Grande do Sul (UFRGS) como objeto de análise. A escolha dessa instituição se justifica, pois está ranqueada como a melhor universidade federal brasileira de 2016 segundo os dados do Índice Geral de Cursos (IGC), que avalia as instituições de ensino superior no país e é realizado pelo Ministério da Educação (MEC) (BRASIL, 2018). As expressões de busca foram testadas em cada uma das bases de dados com as variações do nome da Universidade, sem limitação temporal predeterminada, possibilitando buscar todas as patentes que foram indexadas nas bases. Porém, tendo em vista que nem todas as variações possibilitaram recuperação, as expressões que apresentaram dados pertinentes a pesquisa são detalhadas nos resultados. Não foi utilizado o critério de temporalidade dos registros para a coleta de dados. Por fim, os dados coletados foram registrados em uma planilha Excel para



possibilitar as comparações e análises, bem como observações foram feitas no processo de utilização das bases.

3 RESULTADOS

A pesquisa foi realizada na DII, Orbit, INPI e *Google Patents* com diferentes estratégias de busca. Foram recuperados um total de 1.275 patentes nas quatro bases de dados analisadas, conforme a Tabela 1.

TABELA 1 - RESULTADOS OBTIDOS NAS BASES DE DADOS PESQUISADAS E EXPRESSÕES DE BUSCA UTILIZADAS

Base de Dados	Resultados	Expressões de Buscas
ORBIT	428	(Universidade Federal do Rio Grande do Sul)
DII	410	AN=(UNIV AND FEDERAL AND RIO AND GRANDE AND SUL) OR AN=(UNIV AND FEDERAL AND DO AND RIA AND GRANDE AND DO AND SUL AND UFRGS) OR AN=(UNIV AND FEDERAL AND DO AND RIO AND GRANDE AND DO AND SUL) OR AN=(Hospital AND Clinicas AND PA) OR AN=(Hospital AND Clinicas AND Porto AND Alegre) OR AN=(Agron AND Univ AND Fed AND Rio AND Grande AND do AND Sul) OR AN=(Cenbiot AND UFRGS) OR AN=(Fdn AND Univ AND Fed AND Rio AND Grande AND Sul) OR AN=(Fdn AND Univ AND Fed AND Rio AND Grande AND Sul) OR AN=(Fed AND Univ AND Fed AND Rio AND Grande AND Sul) OR AN=(Fed AND Univ AND Hosp AND Rio AND Grande AND do AND Sul) OR AN=(Fed AND Univ AND Rio AND Grande AND do AND Sul) OR AN=(Fed AND Univ AND Rio AND Grande AND Sul) OR AN=(Fed AND Univ AND Rio AND Grande AND Sul AND State) OR AN=(Fed AND Univ AND Rio AND Grande AND Sul AND UFRGS) OR AN=(Fed AND Univ AND RS) OR AN=(Federal AND University AND Rio AND Grande AND do AND Sul) OR AN=(HCPA) OR AN=(Rio AND Grande AND do AND Sul AND Fed AND Univ) OR AN=(UFRGS) OR AN=(UFRGS) OR AN=(Univ AND Fed AND Rio AND Do AND Sul) OR AN=(Univ AND Fed AND Rio AND Grande AND do AND Sul) OR AN=(Univ AND Fed AND Rio AND Grande AND Sul) OR AN=(URFGS)
Google Patents	33	“UFRGS” OR “Universidade Federal Do Rio Grande Do Sul -UFRGS” OR “Universidade Federal Do Ria Grande Do Sul -UFRGS” OR “Federal University of Rio Grande Do Sul” OR “Univ Federal Do Rio Grande Do Sul - Ufrgs”
INPI	404	UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
TOTAL	1.275	

Fonte: Dados da Pesquisa.



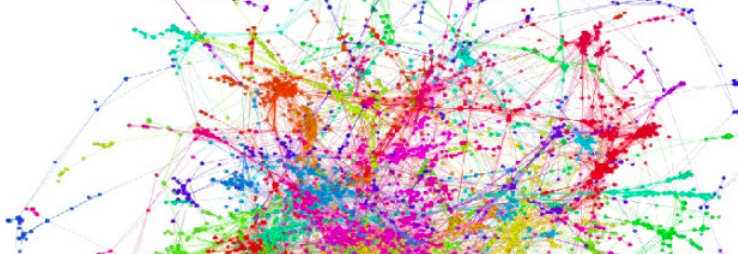
Em relação à cobertura das bases de dados, a única que não indexa patentes registradas no escritório brasileiro (INPI) é a *Google Patents*, pois ela não faz parte dos 17 escritórios mundiais cobertos pela sua busca. Contudo, os resultados obtidos nessa base refletem em quais países/escritórios internacionais a UFRGS também registrou seus pedidos de patentes. De acordo com a Quadro 1, observa-se as diferenças de resultados obtidos e cobertura nas quatro bases de dados:

QUADRO 1 - RESULTADOS RECUPERADOS DII, ORBIT, INPI E GOOGLE PATENTS

	DII	Orbit	INPI	Google Patents
Registros recuperados	410	428	404	33
Registros válidos	390	321	404	33
Patentes com sigla BR	373	318	404	0
Patentes concedidas	390	241	326	33
Patente mais recente nos registros	07/06/2017	21/11/2017	09/01/2018	01/08/2017
Patente mais antiga nos registros	26/08/1997	26/06/1990	04/01/1990	02/09/2005

Fonte: Dados da Pesquisa.

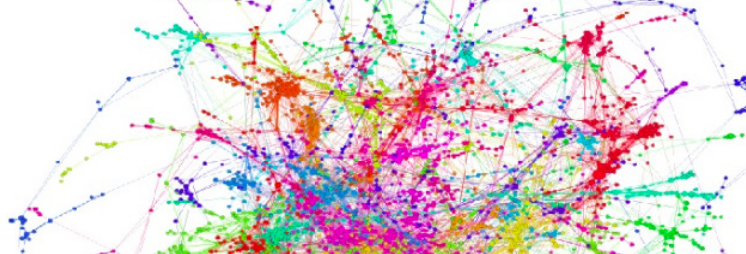
Alguns erros foram identificados nos registros recuperados na DII, por exemplo, a atribuição da Patente BR102016008747-A2 como sendo a depositária a UFRGS, sendo que no INPI a depositária seria a Universidade Federal do Rio Grande do Norte (UFRN). Além disso, a padronização do nome das instituições na DII não é controlada, dentre as variantes do nome da UFRGS foram localizadas também as seguintes variações: “UNIV FEDERAL DO RIO GRANDE DO SUL UFRGS”, “UFRGS INST FISICA”, “UNIV FEDERAL DO RIA GRANDE DO SUL UFRGS (UYRI-Non-standard)” (destaque nosso), “UNIV FEDERAL DO RIO GRANDE DO SUL (UYRI-Non-standard)”, “UNIV FEDERAL RIO GRANDE DO SUL (UYRI-Non-standard)”, “UNIV FEDERAL RIO GRANDE SUL (UYRI-Non-standard)”, “UNIV FED DO RIO GRANDE DO SUL UFRGS (UYFE-Non-standard)” (destaque nosso). Destaca-se ainda, que alguns registros estão com erros



de grafia e outros incorporam a sigla da instituição. Na base de dados Orbit também foram encontradas variações do nome da UFRGS, estando registrada como “UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL” e “UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL - **UFRGS**” (destaque nosso).

Em relação às características dos recursos e dos mecanismos de buscas identificados nas bases de dados, a DII é uma base referencial de patentes que indexa documentos provenientes de mais de 40 escritórios. Além do acesso a mais de 71 milhões de patentes, cobertos desde 1963 (CLARIVATE ANALYTICS, 2017), a indexação feita pela DII acrescenta outras informações ao registro que facilitam a busca e a recuperação de documentos, como, por exemplo, o *Derwent Class Code* (busca e recupera uma determinada categoria de invenção) e o *Derwent Manual Code* (indica novos aspectos técnicos da invenção). As buscas podem ser realizadas por meio de vários rótulos. Os registros recuperados também incluem links para os documentos citados (outras patentes ou artigos de periódicos) atribuídos pelos examinadores das patentes. Todas essas características auxiliam na análise dos dados quantitativos de citação, bem como na busca e recuperação das informações relacionadas à inovação. Ao longo dos anos a DII tem sido utilizada em diversos estudos patentométricos que envolvem a análise dos dados estatísticos de patentes (CONSONI; MOURA, 2016; MARICATO; NORONHA, 2013). Isso se deve principalmente a característica abrangente do seu conteúdo e a existência de mecanismos de exportação dos dados para a posterior análise. Para esta pesquisa, o acesso ao conteúdo ocorreu através do Portal de Periódicos Capes, uma vez que a DII é uma base paga e exige uma assinatura.

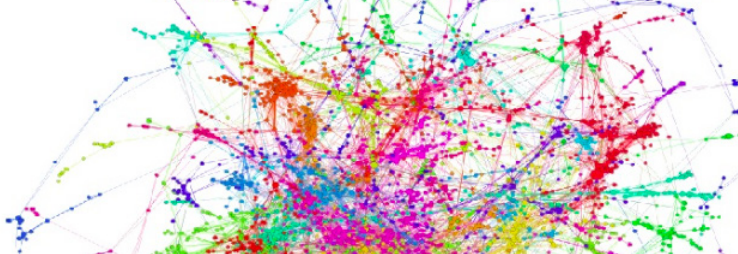
A *Orbit* é uma base de dados paga e disponibilizada por meio de assinatura produzida pela *Questel Orbit* na década de 1970. Oferece um sistema de busca, seleção, análise e exportação de informações contidas em mais de 55 milhões de famílias de patentes e mais de 81 milhões de patentes aplicadas em 100 países e escritórios de registros de patentes. Ela apresenta dados sobre depositantes, inventores, textos com traduções para diversos idiomas, imagens e citações, reunidas em ba-



ses de dados proprietárias: *FamPat* e *FullPat* (respectivamente família de patentes e pedidos de patentes). As principais funcionalidades da base podem ser assim agrupadas: busca, visualização, seleção, análise, exportação, monitoramento e compartilhamento. Ela possui uma tabela família de patentes que permite personalizar e exibir valores métricos para cada família em seu conjunto de análise. É possível escolher as métricas mais relevantes para uma pesquisa, classificar por valores de métricas para identificar as patentes-chave e exportar o conjunto de dados completo em vários formatos de arquivos, possibilitando, assim, realizar as análises detalhadas.

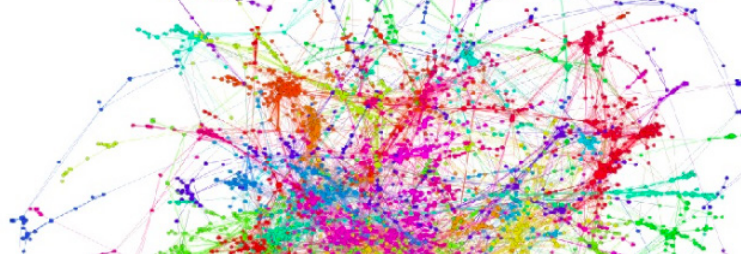
O INPI disponibiliza em seu site um recurso de busca onde é possível pesquisar as patentes por diferentes campos, podendo mesclá-los, mas não permitindo cruzar as informações ou aplicar estratégias de buscas mais elaboradas. Nesse recurso, não há possibilidade de exportação dos dados, dificultando a recuperação de grandes quantidades de registros. Por isso, desde agosto de 2017, de forma a se adequar ao Decreto nº 8.777/2016, o INPI passou a disponibilizar no Portal de Dados Abertos a Revista de Propriedade Intelectual (RPI), com a possibilidade de baixar os metadados das seções da revista, contendo informações sobre o registro de solicitações, recursos e concessões de patentes. A RPI encontra-se disponível *on-line* e é publicada semanalmente às terças-feiras pelo INPI, em uma numeração sequencial e incremental. A revista está estruturada em oito seções: I) comunicados; II) contratos de tecnologia; III) desenhos industriais; IV) indicações geográficas; V) marcas; VI) patentes; VII) programas de computador e; VIII) topografia de circuito integrado (BRASIL, 2017). Para fins dessa pesquisa foram baixados os arquivos da RPI de P1132.ZIP, edição de 11/08/1992 até o P2450.ZIP, edição de 19/12/2017. Esse recurso permite identificar o nome e endereço URL de todas as RPIs disponíveis em formato PDF, ZIP+TXT¹ e/ou ZIP+XML até 24/Jan/2017.

O *Google Patents* é uma base de acesso gratuito que permite a busca e leitura de textos completos de patentes de vários escritórios de depósitos de patentes. Disponibiliza um conjunto de mais de 90 milhões de



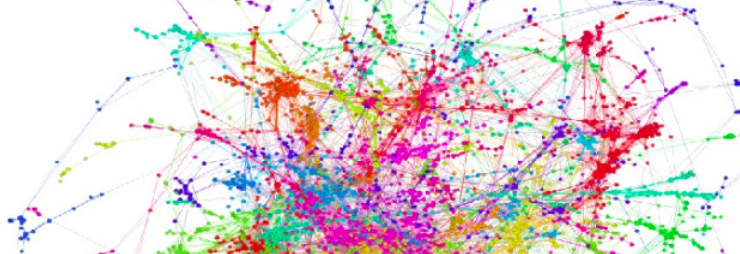
patentes de 17 escritórios mundiais de registros de patentes, tornando esses dados públicos por meio do *Big Query*, que é uma ferramenta de armazenamento de dados (WETHERBEE, 2017). A interface de busca é simplificada, similar a busca do *Google*. Além disso, a base disponibiliza a ferramenta *Search and read the full text of patents from around the world* que permite visualizar em tempo real a quantidade de patentes concedidas e aplicadas em cada um dos 17 escritórios cobertos. Oferece, também, o recurso de pesquisa em documentos no *Google Scholar* (*Include non-patent literature*), permitindo a busca de anterioridade e estado da arte daquilo que se deseja patentear. O *Google Patents*, por meio de classificação automática, utiliza a Classificação Cooperativa de Patentes para que os resultados sejam recuperados (GOOGLE, 2018).

O Quadro 2 apresenta as principais características da recuperação e padronização dos dados nas diferentes bases de dados analisadas.

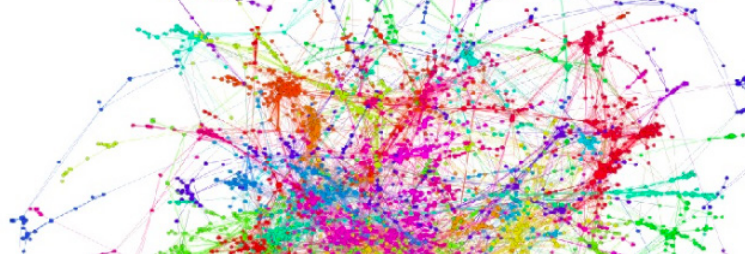


QUADRO 2 – CARACTERÍSTICAS DA RECUPERAÇÃO E PADRONIZAÇÃO DOS DADOS DII, ORBIT, INPI E GOOGLE PATENTS

Língua	DII	Orbit	INPI	Google Patents
	Interface em português. Descrições dos registros em inglês	Interface em inglês.	Interface em português	Interface em inglês
Apresentação dos resultados	Número da patente, título, inventores, depositantes e número de acesso primário na DII, Quando a patente está disponibilizada na íntegra, é possível baixar o documento clicando no botão “original”. Possibilita a visualização de até 50 registros por página	Número de família de patente, número prioritário, título, inventores, depositantes, PDF do documento original, imagens, resultado dos dados, recursos de leitura e seleção de resultados, associadas a ferramentas de exportação de dados, com hyperlinks para textos integrais, citações, palavras-chaves no contexto (KWIC), ferramenta de <i>highlight</i> (colorir termos de busca), gráficos de citações, gráficos de famílias de patentes, ferramenta de comparação de reivindicações, ferramenta de reordenamento dos resultados.	Número da patente, data de solicitação, título da patente e sua classificação. Para acesso aos dados completos é necessário acessar o número da patente. Visualização de 20 registros por página.	Número da CIP, título, dados do inventor e depositante, datas de depósito, publicação e indexação na base, pequena parte do resumo, possibilidade de realizar outra busca com o mesmo número de CIP da patente recuperada. Possibilita a visualização de até 100 registros por página
Forma de exportação dos dados	De 500 em 500 registros	Permite exportar em diversos formatos, sem limite no número de exportação. Além disso, permite exportar: texto, imagem, campos selecionados, Hyperlinks permanentes (imagens, textos integrais pdf, citações, dados da família de patentes)	No site de busca não existe exportação de dados. A partir de agosto de 2017 os dados da RPI estão no portal Dados Abertos do Governo Federal com todas suas seções.	Exporta todos os resultados em uma única vez
Formato de exportação dos dados	html ou .txt, com possibilidade de conversão do arquivo .txt para Excel, tanto para Windows como para Macintosh	.txt, .pdf, .xls, .xlsx, .csv, .xml	Não aplicado. Para o portal de dados abertos, os dados estão em .txt	.csv



Informações contidas na exportação	Número da patente (PN), título (TI), inventor (AU), depositante (AN), data de indexação (GA), resumo (AB), resumo expandido (TF), código de classe da DII (DC), código manual da DII (MC), CIP (IP), ano de publicação (PD), ano de depósito (AD), patentes citadas (CP), artigos citados (CR)	Título (TI), Resumo (AB), Classificação Internacional ou Classificação Europeia ou CPC (IC/EC/CPC), Basic Index (BI) Super-Abstract (SA), número de publicação (PN), Número de publicação de família (NPN), Número Prioritário de família (NPR), Data da primeira publicação da família de patentes (EPD)	(21) N° do Pedido; (22) Datado Depósito; (43) Data da Publicação; (47) Data da Concessão; (51) Classificação IPC; (54) Título; (57) Resumo; (71) Nome do Depositante; (72) Nome do Inventor; (74) Nome do Procurador	Número da patente (ID), título (Title), depositante (Assignee), inventor (Inventor /author), data de depósito (Priority date), data de indexação (Filing/creation date), data de publicação (Publication date), data de concessão (Grant date) e o link para o registro na base
Campos de busca	Busca simples e avançada com os campos: Assunto, título, inventor, depositante, número da patente, CIP, código de classe na DII, código manual na DII, número de acesso primário na DII, código do depositante, nome mais código do depositante, número de patente citada	Apresenta diferentes formas de busca, tais como: busca básica, busca por menus, busca por linhas de comando avançadas, busca por citações, busca por similaridade, busca por famílias de patentes expandidas, assistente de busca semântica, assistente de busca por números de publicação, assistente de busca por titulares, assistentes de busca por classificações, buscas por conjuntos de resultados previamente selecionados.	Busca simples e avançada com os campos: Número do pedido, número do recolhimento, do protocolo, data de depósito, data de prioridade, data de publicação, número de classificação, palavra-chave pelo número de classificação, palavra-chave no título e no resumo, nome do depositante, CPF ou CNPJ do depositante e nome do inventor	Busca simples e avançada com os campos: termos, data de depósito, inventor, depositante, escritório, língua, status e tipo
Disponibilidade de filtros na base	Sim	Sim	Não	Não
Dados de citação	Links para as citações dos examinadores	Hyperlinks permanentes de citações.	Não	Dados das patentes que foram citadas no documento e das patentes que já citaram a patente que está sendo visualizada

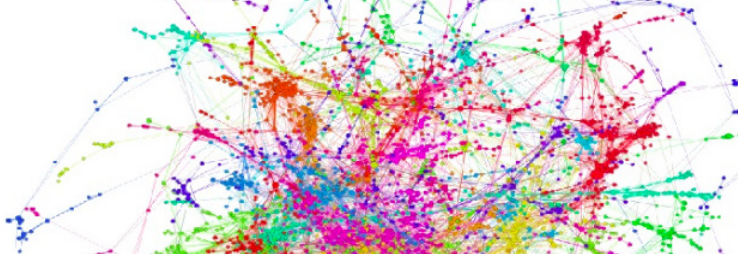


Recursos de análise dos resultados	Recurso “Análise de Resultados” com contagem do número de registros por depositantes, inventores, CIPs e área do conhecimento	Contagem do número de depositantes, inventores e CIPs do resultado da pesquisa. Apresenta gráficos, grafos, tabelas, análise da evolução temporal das subclasses internacionais, mapas geográficos.	Não	Contagem do número de depositantes, inventores e CIPs do resultado da pesquisa. Mostra um gráfico com a contagem relativa dos 5 primeiros resultados em relação ao ano de indexação na base
Cobertura de tempo	1963 - presente	1960 - presente	1970 - presente	1790 - presente

Fonte: Dados da Pesquisa.

4 CONSIDERAÇÕES FINAIS

Os estudos de patentes ou patentométricos estão sendo cada vez mais comuns na academia, buscando a compreensão de como se apresentam as relações entre a ciência e a tecnologia. Nesse contexto, este estudo teve como objetivo analisar e comparar algumas fontes de informações que disponibilizam dados sobre as patentes, possibilitando melhor entendimento das características de cada uma, sua abrangência e singularidades. Para esse fim, utilizou-se a UFRGS como elemento de análise nas bases para realização das comparações. Observou-se que as características das bases são diferentes, seja por sua abrangência, escritórios que indexa, formas de recuperação, de exportação e tempo de atualização. Considerando o objeto do estudo, as patentes depositadas no Brasil, o INPI apresenta a melhor cobertura. No entanto, se a necessidade é analisar as relações em diversos países, a DII e a *Orbit* podem ser mais eficientes, pois indexam patentes de vários escritórios mundiais. Entretanto, observou-se que na DII há falta de padronização no nome de inventores e depositantes, não somente na grafia, mas também nos registros de forma trocada de nomes das instituições depositantes. A *Orbit* é a que tem melhor qualidade dos dados, apresentando atualização periódica e disponibilização de análises detalhadas dos resultados, entretanto é uma base paga. O *Google Patents* foi o que apresentou piores resultados para o caso da UFRGS, uma vez que não indexa as paten-



tes registradas no escritório brasileiro, necessitando mais análises com outras instituições. Dessa forma, pode-se concluir que nos estudos de patentes a escolha da fonte de dados é muito importante para a qualidade do estudo, devendo-se escolher conforme o objeto a ser estudado, a abrangência, atualidade e possibilidades de recuperação e exportação.

REFERÊNCIAS

BRASIL. Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Índice Geral de Cursos (IGC)**. Brasília, DF, 2018. Disponível em: <<http://portal.inep.gov.br/indice-geral-de-cursos-igc->>. Acesso em: 19 abr. 2018.

_____. Ministério do Planejamento, Desenvolvimento e Gestão. Secretaria de Tecnologia da Informação. **RPI de Patentes (a partir de 31/Jan/2017)**. Brasília: Portal Brasileiro de Dados Abertos, 2017.

CALDAS, M. S.; SILVA, E. C. C. Fundamentos e aplicação do Big Data: como tratar informações em uma sociedade de yottabytes. **Revista Bibliotecas Universitárias: Pesquisa, Extensão e Perspectiva**, v. 3, n. 1, p. 65-85, jan./jun. 2016.

CLARIVATE ANALYTICS. **Derwent World Patents Index**. [S.l.]: Clarivate Analytics, 2017.

CONSONI, L. A. E. A.; MOURA, A. M. M. Patentes de biodiesel indexadas na derwentinnovations index entre 2008 e 2009: primeiras considerações. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 17., Salvador, 2016. **Anais...** Salvador: ANCIB, 2016.

GOOGLE. **About Google Patents**. [S.l.]: Google, 2018. Disponível em: <<https://support.google.com/faqs/answer/6390996>>. Acesso em: 04 jan. 2018.

MARICATO, J. O. M.; NORONHA, D. P. Coclassificação em artigos e patentes em biodiesel. **Liinc emRevista**, v. 9, n. 1, p. 85-102, 2013.

WETHERBEE, I. **Google Patents Public Datasets**: connecting public, paid, and private patent data. [S.l.]: Google, 2017.